

## Comparación de métodos predictivos en el estudio de la distribución potencial de especies forestales

M. Anaya-Romero, A. Jordán López, L. Martínez Zavala & N. Bellinfante Crocci

*Dpto. de Cristalografía, Mineralogía y Química Agrícola. Facultad de Química (Univ. de Sevilla). C/Profesor García González, 1. CP: 41012 – Sevilla (España)*

En este trabajo se ha realizado una comparación entre diversos modelos predictivos de la vegetación forestal en áreas naturales del Parque Natural Sierra de Aracena y en el Andévalo Occidental (Huelva, España). Los modelos comparados han sido la regresión logística, la red neuronal artificial y los árboles de decisión. En todos ellos, la predicción se ha hecho relacionando la presencia/ausencia de cada formación forestal con variables edáficas, climáticas y geomorfológicas que, a priori, pudieran condicionar la distribución de las especies.

En el área se han diferenciado cuatro formaciones forestales principales: quercíneas, coníferas, eucaliptos y otras frondosas.

Se realizó la clasificación de las formaciones forestales presentes en la zona en cuatro grupos resultaron ser “quercíneas”, “coníferas”, “eucaliptos” y “otras frondosas”. Posteriormente se analizó la distribución actual de cada una de estas formaciones en el área de estudio.

En cuanto a las variables ambientales seleccionadas, estas se agruparon en varias categorías: litología (tipo de roca, acidez de la roca y consolidación de la roca), procesos geomorfológicos (procesos erosivos, movimientos en masa, sedimentación y morfogénesis), clases fisiográficas, topografía (elevación, pendiente, curvatura, y orientación de la ladera), variables edáficas (propiedades físicas y químicas del horizonte superficial del perfil del suelo: acidez del suelo, oligoelementos asimilables, macroelementos asimilables, fósforo asimilable, materia orgánica, capacidad de intercambio catiónico, saturación del complejo de cambio, porcentaje de gravas, y porcentaje de arcilla en la fracción de tierra fina) y climatología (precipitación anual y media en verano, temperatura media anual, temperatura media del mes más caluroso y temperatura media del mes más frío). Para cada una de estas variables se realizó un mapa de distribución (presencia/ausencia) o de clases.

Previamente a la aplicación de los tres modelos de evaluación, se realizó el procesamiento de la información para poder integrarla de una forma homogénea en una base de datos georeferenciada.

Tras el análisis exploratorio de los datos, se realizó un muestreo uniforme sobre el que se aplicaron los tres modelos de predicción, relacionando los usos forestales con las variables ambientales.

Con el objetivo de estudiar el comportamiento de cada análisis de predicción utilizado frente a datos nuevos, se dividió el total de la información en datos de entrenamiento y datos de test. La división se realizó de forma aleatoria, de manera que el 75% de los valores pertenecen a datos de entrenamiento, y el 25% restante corresponde a los valores de test. Los resultados obtenidos con los tres modelos de predicción se presentaron en forma de matriz de confusión. En dicha matriz se comparan los valores reales con los valores predichos por los distintos métodos. Para comparar la idoneidad en cada uso forestal de los métodos empleados, se calculó un índice de error de estimación a través de la matriz, que consiste en la suma de los errores predichos para la presencia estimada dividida por la suma total de puntos analizados. Dicho error de estimación se calculó tanto para los valores de entrenamiento como para los valores de test, siendo más relevantes a la hora de evaluar el modelo los resultados obtenidos en los valores de test.

En general, se observó, que los modelos de regresión logística y árbol de decisión ofrecen resultados similares, y ambos poseen un índice de error menor que la red neuronal artificial en las cuatro formaciones forestales. Finalmente se realizó el promedio del índice de error de estimación en las cuatro formaciones forestales para cada modelo de evaluación utilizado. La regresión logística presentó mejores resultados en los datos de tipo test, por lo que se seleccionó para el desarrollo de un modelo de distribución potencial de los cuatro usos forestales presentes en la zona de estudio. Los resultados se extrapolaron a la totalidad de las celdas del territorio, de esta forma se obtuvo el hábitat potencial de las formaciones forestales estudiadas.